Spring 2020

CSE 351 : Introduction to Data Science
CSE 519 : Data Science Fundamentals

LECTURE 0 – COURSE INTRODUCTION

1

Course Information

CSE351/CSE519 : Introduction to Data Science

Course webpage: http://ppawar.github.io/Spring2020/CSE351-S20/index.html

Lectures: Mon/Wed 10:30 AM - 11:50 AM

Place: Online Zoom meetings will be scheduled for first two weeks

2

Staff

Instructor

- Pravin Pawar
- Office: B424
- Email: Pravin.pawar@sunykorea.ac.kr
- Phone: +82-032-626-1227 / +82-010-8692-4908
- Office Hours: *Tue/Thu 10:30 AM - 12:30 PM*
- *(Will be available on Skype, Kakao talk and email. Skype ID is pravin.pawar)*

3

Announcements

Zoom meeting invitation will be sent in advance for the specific class times.

The zoom meeting session will be recorded and will be made available for viewing later online.

It is expected that you attend each lecture online (unless medical situation).

The instructor will record your attendance in-between the lecture break on blackboard.

Please bring a laptop to each class

- Classes will involve lecture segments, demos
- Labs will involve student exercises

Additional video lectures are noted in the syllabus. These are strongly recommended for extra instruction to help understand various technologies we will learn in this course.

4

Course Overview

- This multidisciplinary course introduces both theoretical concepts and practical approaches to extract knowledge from data.
- Topics include linear algebra, probability, statistics, machine learning, and programming.
- Using large data sets collected from real-world problems in areas of science, technology, and medicine, we introduce how to preprocess data, identify the best model that describes the data, make predictions, evaluate the results, and finally report the results using proper visualization methods.
- This course also teaches state-of-the art tools for data analysis, such as Python and its scientific libraries.

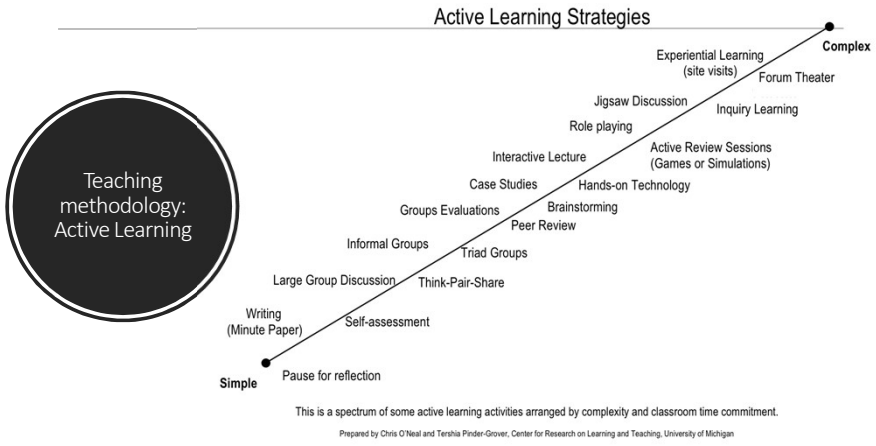
5

Major Course Topics

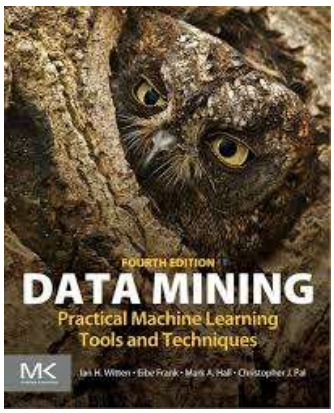
- Introduction to data science
- Relationship between data science and machine learning
- Mathematical and statistical preliminaries
- Introduction to machine learning algorithms
- Handson with data mining software (one of the following: Weka, Orange, KNIME, RapidMiner)
- Data science with Python packages such as NumPy, Pandas
- Data visualization
- Regression
- Classification
- Clustering
- Associative rule mining
- Genetic algorithms
- Introduction to neural networks and deep learning
- A handson group project

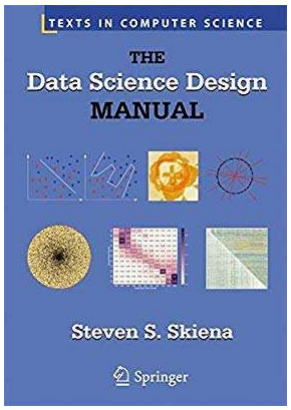
6

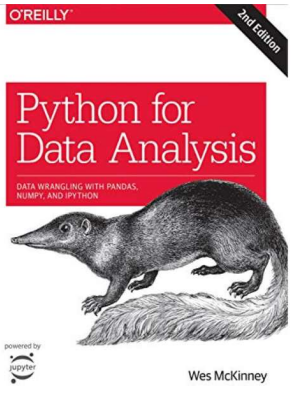
Teaching methodology: Active Learning



7

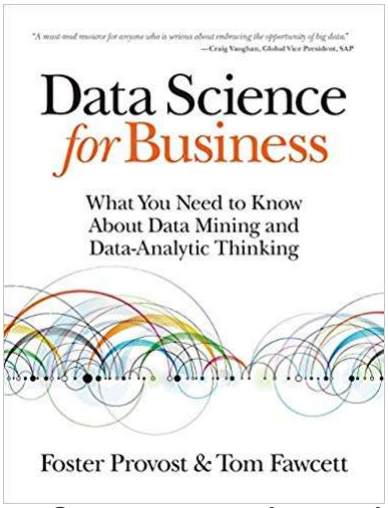


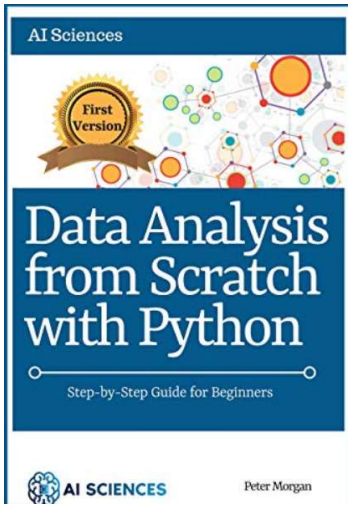




Textbooks

8





Reference books

9

Homework Assignments

- These homework assignments will reinforce concepts from class and have you explore new concepts, too
- All work will due on fixed dates and times
- All work will be completed on an individual basis (write your own code) *unless otherwise instructed*!
- You will use **Blackboard** to submit your completed assignments
- Please start early on the assignments! Most students find that completing the homework assignments for CSE351/CSE519 takes **a lot** longer than they anticipated

10

Examinations

- Examination dates are posted on the schedule page of the course website. Tentative dates are:
  - Exercises/Quizzes: Each Wednesday, there will be either classroom exercise or quiz
  - Midterm exam 1: Mon 13 Apr
  - Midterm exam 2: Mon 18 May
  - Final exam: Wed 17 June, 9:00 AM – 11:30 AM – Datathon
- Do not miss exams
- Arrange your work and travel schedules as needed to be present for examinations
- Makeup exams will only be given for verified, officially sanctioned university activities.
- All midterms will be closed-notes and closed-book, except one sheet of notes (A4 or 8.5x11), both sides (handwritten or typed)

11

Grading

- Assignments – Handson data analytics assignments (30%, 4 assignments given) = 30% (150 points).
- Class exercises and surprise quizzes (~10, 2% each) = 20% (100 points).
  - A number of class exercises and quizzes will given during the lecture on the material covered in class that week. These exercises and quizzes are intended to make sure students are keeping up with the material.
- Mid-term exams (30%, 2 exams given) = 30% (150 points).
  - These will be written exams.
- Live Final exam = 20% (100 points).
  - For this exam, you will be given a data analytics challenge to solve.
- Policies:
  - Makeup exams will only be given for verified, officially sanctioned university activities

12

Late Homework Policy

- Assignments must be turned in by the due date and time.
  - Any part of an assignment that's late means the entire assignment is late.
  - If your assignment is incomplete or not entirely working by the due date, turn in what you have to get some partial credit.
- If you have an emergency situation, email me before the due date and I may be able to work something out
- Bottom line: Plan ahead, start early!

13

Re-Grading

- For the assignments, quizzes and mid-term exams, request for re-grading must be made within one week from after the announcement of grades.

14

Pair Programming

- Identify your programming buddy
- Sit next to each other in a class as much as possible
- Discuss your problems with him/her
- Solve exercises available online and from the books together
- Help each other to learn the course in a cooperative way!!

Benefits:

- Fewer bugs
- Spreads code understanding
- Higher quality code
- Can learn from partner
- Two heads are better than one
- Creativity and brainstorming
- Better testing and debugging
- Improved morale

15

Cooperation vs. Copying

- Cooperation (talking over problems) is a good way to learn and is encouraged
- ***Do not copy code. Do not let others look at or copy your code.***
- Copying is not allowed on homework or exams no matter the source
- When you submit your homework or tests, **you are pledging that the work is your own and you have not copied it.**
  - **You are also pledging that you have not allowed others to copy it**.
- **DO NOT COPY! (Our grading TA and *software tools catch it easily*)**

16

Electronics in Class

- Cell phones should be put away during class
- Laptops may be used during periods where you are asked to work on an exercise during class
- Lecture slides are available on the course website for study before class
- Talk to me after class if there's an issue with this policy

17

Disability

If you have a physical, psychological, medical or learning disability, please contact the Student Services and Career Team.

- Location: Academic Building A208
- Phone: 626-1190

The DSS will determine with you what accommodations, if any, are necessary and appropriate

All information and documentation of disability is confidential

18

How to Succeed in this Class

Attend class and be on time!

- Not all information is in my lecture notes or in the book
- I sometimes do in-class demos that emphasize non-obvious details

This is an introductory course, true, but we're going to cover a lot of ground and move quickly starting from scratch

The assigned work will take a lot of your time, so practice good time management

Read the reading assignments and review the lecture notes and try out example code

- Practice is the only way to become proficient at coding
- Very often your first, second, or third attempt at solving a problem will not be successful. It is **essential** that you give yourself enough time to try different ideas, taking breaks along the way!
- Those who write extra code for problems not assigned ("for fun") generally do best in this class
- Learning to code involves learning to read other people's code

Ask questions right away if confused. Ask in class, ask a TA, come to my office hours or send email. Don't stay confused and don't get behind!

Welcome and I hope you enjoy the class!

19

Questions?

Online discussion forum: Piazza

Find our class page at: https://piazza.com/sunykorea.ac.kr/spring2020/CSE351/home

Maintain decorum, take responsibility, no anonymous questions.

You will be enrolled in Piazza.

Tutorial: https://rutgers.instructure.com/courses/35/pages/piazza

20